

Title: Consistent Estimation of the Number of Communities via Regularized Network Embedding

Authors: Mingyang Ren, Sanguo Zhang, Junhui Wang

Code maintainer: Mingyang Ren (renmingyang17@mails.ucas.ac.cn)

R version: 4.0.5 (2021-03-31) (platform: x86_64-w64-mingw32/x64 (64-bit))

R package versions:

RSpectra_0.16-0 pracma_2.3.8 randnet_0.5 AUC_0.3.0 entropy_1.3.1

Matrix_1.3-2 MASS_7.3-53.1

R files:

The following .R files are included in the submitted supplementary files.

(Main functions for the implementation of the proposed method)

1. `./main_functions/function.R`: This file includes all main functions of proposed methods to support numerical simulation studies and real data analysis.
2. `./main_functions/alternatives.R`: This file includes alternative methods to support numerical simulation studies and real data analysis.

(Codes for simulation studies)

3. `./simulations/sim-func.R`: This file includes functions for generating simulated data and evaluating performances of competitors, which are used to support numerical simulation studies.
4. `./simulations/sim_func_para.R`: This file includes parameter setting functions required for simulations.
5. `./simulations/main-simulation.R`: This file is used to conduct simulation examples.

(Codes for real data analysis)

6. `./case_study/data-process.R`: This file is used for data preprocessing.
7. `./case_study/community_detection.R`: This file is used for community detection of ADHD brain network.
8. `./case_study/data.analysis.main.R`: The main program applying the proposed method to perform ADHD brain network analysis.

How to conduct simulation examples:

Run `./simulations/main-simulation.R`

How to conduct real data analysis:

--0. Data preparation:

- (1) "Node_AAL116_info.csv" in working path "`./case_study/`"
- (2) AAL-NYU Time Courses data, "NYU_phenotypic.csv", and "NYU-group.csv" files in subfolder "`./case_study/data`", which are downloaded from https://www.nitrc.org/plugins/mwiki/index.php/neurobureau:AthenaPipeline#Whole_Brain_Data

--1. Data preprocessing:

Run "data-process.R", and output the adjacency matrix of two subtypes, which is saved as "correlation-AAL-NYU.RData"

--2. Data analysis:

Run "data.analysis.main.R", applying the proposed method for community detection,

(1) outputting the figure of adjacency matrix of ADHD brain functional connectivity networks and, i.e., "ADHD.adjacency.com.eps" and "ADHD.adjacency.ina.eps".

(2) outputting the .node and .edge files needed to draw the brain network figures in the path ".case_study/results".